

**PENGELOMPOKAN HASIL TANGKAPAN NELAYAN KABUPATEN CILACAP
MENGUNAKAN METODE *K-MEANS CLUSTERING***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer pada
Departemen Ilmu Komputer/ Informatika**

**Disusun Oleh:
IMAN SETYA ADJI
24010313140102**

**DEPARTEMEN ILMU KOMPUTER/INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini :

Nama : Iman Setya Adji

NIM : 24010313140102

Judul : Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap Menggunakan Metode *K-means Clustering*

Dengan ini menyatakan bahwa dalam tugas akhir / skripsi ini tidak terdapat karya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 7 Agustus 2019



Iman Setya Adji

NIM. 24010313140102

HALAMAN PENGESAHAN

Judul : Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap Menggunakan Metode *K-means Clustering*
Nama : Iman Setya Adji
NIM : 24010313140102

Telah diujikan pada sidang tugas akhir dan dinyatakan lulus pada 7 Agustus 2019.

Mengetahui,

a.n Ketua Departemen Ilmu

Komputer/Informatika

Sekretaris



Dr. Eng. Adi Wibowo, S.Si, M.Kom.

NIP. 198203092006041002

Panitia Penguji Skripsi,

Ketua,

Dr. Retno Kusumaningrum, S.Si, M.Kom

NIP. 198104202005012001

HALAMAN PENGESAHAN

Judul : Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap Menggunakan Metode *K-means Clustering*

Nama : Iman Setya Adji

NIM : 24010313140102

Telah diujikan pada sidang tugas akhir pada 7 Agustus 2019.

Semarang, 7 Agustus 2019

Pembimbing,



Sukmawati Nur Endah, S.Si, M.Kom

NIP. 197805022005012002

ABSTRAK

Kabupaten Cilacap sebagai Kabupaten yang memiliki wilayah pantai merupakan salah satu daerah penghasil produk laut yang ada di Indonesia. Koperasi Unit Desa Mino Saroyo melayani nelayan dalam proses pemasaran produk laut dengan mengelola Tempat Pelelangan Ikan yang terdapat di Kabupaten Cilacap. Dalam tugasnya untuk mengelola Tempat Pelelangan Ikan, Koperasi Unit Desa Mino Saroyo mencatat setiap lelang yang dilaksanakan. Data yang dicatat hanya digunakan sebagai laporan hasil lelang. Penelitian ini menerapkan *data mining* untuk memperoleh pola dan informasi baru dari data tangkapan nelayan. Pengolahan data laporan hasil lelang koperasi menjadi data yang layak digunakan untuk data mining merupakan bagian dari penelitian ini. Untuk mencari informasi baru, metode *K-means Clustering* digunakan untuk mengelompokkan data. Informasi baru yang dicari berupa pola unik yang dihasilkan dari hasil pengelompokan dan pengujian hubungan antar atribut data yang kemudian ditampilkan menggunakan visualisasi data. Hasil penelitian diperoleh informasi kelompok data tangkapan nelayan berdasarkan produksi dan harganya. Selain itu diketahui bahwa hasil pengelompokan data ternyata memiliki hubungan dengan nilai jenis tangkapan.

Kata Kunci : Tangkapan Nelayan, *Data Mining*, *K-means Clustering*

ABSTRACT

Cilacap Regency as a Regency which has a coastal area is one of the producing areas of marine products in Indonesia. Koperasi Unit Desa Mino Saroyo serves fishermen in the process of marketing marine products by managing the Fish Auction Place in Cilacap Regency. In their task of managing the Fish Auction Place, Koperasi Unit Desa Mino Saroyo records every auction conducted. The data recorded is only used as a report on auction results. This study applies data mining to obtain new patterns and information from fishermen's catch data. Data processing of cooperative auction report data into data that is suitable for data mining is part of this research. To find new information, K-means Clustering method is used to classify data. New information is sought in the form of unique patterns that result from grouping and testing the relationships between data attributes which are then displayed using data visualization. The results obtained by the group of fishermen catch data information based on production and price. In addition it is known that the results of grouping the data turned out to have a relationship with the value of the type of catch.

Keyword : Fishermen catches, Data Mining, K-means Clustering

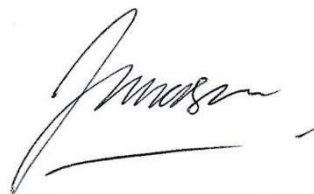
KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasihNya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap Menggunakan Metode *K-means Clustering*” dengan baik dan lancar. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang. Dalam penyusunan tugas akhir ini penulis banyak mendapat bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terimakasih kepada:

1. Prof. Dr. Widowati, M.Si, selaku Dekan FSM UNDIP
2. Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Ketua Departemen Ilmu Komputer/ Informatika
3. Panji Wisnu Wirawan, S.T, M.T, selaku Koordinator Tugas Akhir
4. Sukmawati Nur Endah, S.Si, M.Kom, selaku Dosen Pembimbing

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari penyampaian materi maupun isi dari materi itu sendiri. Hal ini dikarenakan keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan. Semoga laporan tugas akhir ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, 7 Agustus 2019



Iman Setya Adji

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai Civitas Akademik Universitas Diponegoro, saya yang bertanda tangan dibawah ini :

Nama : Iman Setya Adji
NIM : 24010313140102
Program Studi : Informatika
Departemen : Ilmu Komputer/ Informatika
Fakultas : Sains dan Matematika
Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Non-eksklusif** (*Non-exclusive Royalty Free Right*) kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap Menggunakan Metode K-means Clustering

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini, Universitas Diponegoro berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasika tugas akhir ini tanpa meminta izin dari saya selama mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 7 Agustus 2019

Yang menyatakan



Iman Setya Adji

NIM. 24010313140102

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	i
HALAMAN PENGESAHAN	ii
HALAMAN PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup	4
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian Terkait.....	6
2.2. <i>Data mining</i>	8
2.3. Visualisasi Data	10
2.4. <i>Cluster analysis</i>	12
2.5. <i>K-means Clustering</i>	13
2.6. Hubungan Antar Data	14
2.6.1. Korelasi Pearson	15

2.6.2.	Korelasi Spearman.....	16
2.6.3.	Koefisien Kontingensi	16
2.6.4.	Koefisien Cramer.....	18
2.7.	<i>Sum of Squared Error</i>	19
BAB III METODOLOGI PENELITIAN		21
3.1.	Lokasi Penelitian.....	21
3.2.	Tahapan Penyelesaian Masalah	21
3.3.	Pengumpulan dan Pengenalan Data.....	21
3.4.	<i>Data preprocessing</i>	23
3.4.1.	<i>Data cleaning</i>	26
3.4.2.	<i>Data integration</i>	29
3.4.3.	<i>Data transformation</i>	30
3.4.4.	<i>Data selection</i>	34
3.5.	<i>Data Mining</i>	34
3.5.2.	Penentuan nilai k dan inisialisasi <i>centroid</i>	35
3.5.3.	<i>Clustering</i> data.....	35
3.6.	<i>Pattern Evaluation</i>	40
3.6.2.	Evaluasi hasil <i>Clustering</i>	41
3.6.3.	Evaluasi Hubungan Hasil <i>Clustering</i> Terhadap Atribut Lain	42
3.7.	<i>Knowledge Presentation</i>	42
BAB IV HASIL DAN PEMBAHASAN.....		44
4.1.	Lingkungan Penelitian	44
4.2.	Pengujian	45
4.2.1.	Pengujian <i>Clustering</i>	45
4.2.2.	Pengujian Hubungan Antar Data.....	48
4.3.	Hasil Visualisasi Data	53
4.3.1.	Visualisasi Hubungan Antara <i>Cluster</i> Berdasarkan Jenis	54

4.3.2.	Visualisasi Hubungan Antara <i>Cluster</i> dengan Bulan	57
4.3.3.	Visualisasi Hubungan Antara <i>Cluster</i> dengan Tahun	58
BAB V Kesimpulan dan saran		61
5.1.	Kesimpulan	61
5.2.	Saran	61
DAFTAR PUSTAKA.....		62
LAMPIRAN – LAMPIRAN		64

DAFTAR GAMBAR

Gambar 2.1 Langkah – langkah <i>data mining</i> secara umum (Han, et al., 2012).	10
Gambar 2.2 Contoh persebaran data pada <i>Scatter Plot</i>	11
Gambar 2.3 Contoh perbandingan nilai menggunakan <i>Bar Chart</i>	11
Gambar 2.4 Contoh perbandingan rasio menggunakan <i>Pie Chart</i>	12
Gambar 2.5 Grafik hubungan antar variabel	15
Gambar 2.6 Hubungan antara nilai k dengan SSE (Thinsungnoena, et al., 2015).	20
Gambar 3.1 Tahapan penyelesaian masalah	22
Gambar 3.2 Data hasil tangkapan nelayan pada aplikasi Microsoft Excel	23
Gambar 3.3 Data yang dipilih dari berkas spreadsheet	24
Gambar 3.4 Data hasil tangkapan nelayan Kabupaten Cilacap dengan format csv	25
Gambar 3.5 Perubahan format berkas format xls menjadi csv	25
Gambar 3.6 Gambaran penanganan data pada data cleaning	26
Gambar 3.7 Proses dalam tahapan <i>data cleaning</i>	26
Gambar 3.8 Gambaran penanganan data pada <i>data integration</i>	29
Gambar 3.9 Proses dalam tahapan <i>data integration</i>	29
Gambar 3.10 Gambaran penanganan data pada <i>data transformation</i>	30
Gambar 3.11 Gambaran penanganan data pada <i>data selection</i>	34
Gambar 3.12 Gambaran penanganan data pada <i>data mining</i>	35
Gambar 3.13 <i>Flowchart</i> untuk proses <i>K-means Clustering</i>	36
Gambar 3.14 Gambaran penanganan data pada pattern evaluation	41
Gambar 3.15 Dataset yang diekspor dalam format csv	42
Gambar 3.16 Visualisasi yang dibuat di tahap <i>knowledge presentation</i>	43
Gambar 4.1 Daftar percobaan <i>clustering</i> pada data	45
Gambar 4.2 Pengaruh nilai k terhadap nilai SSE	46
Gambar 4.3 Persebaran data berdasarkan berat dan harga perkilogram	47
Gambar 4.4 Tabel Kontingensi dari atribut jenis dan kluster	49
Gambar 4.5 Koefisien korelasi atribut kluster dengan jenis	49
Gambar 4.6 Tabel kontingensi dari atribut kluster dan bulan	51
Gambar 4.7 Koefisien korelasi dari atribut kluster dan bulan	51
Gambar 4.8 Tabel kontingensi dari atribut kluster dan tahun	52
Gambar 4.9 Koefisien korelasi dari atribut kluster dan tahun	52

Gambar 4.10 Implementasi visualisasi distribusi data	53
Gambar 4.11 Implementasi visualisasi Distribusi produksi / penjualan	54
Gambar 4.12 Implementasi visualisasi distribusi jenis tiap <i>cluster</i>	54
Gambar 4.13 Hubungan antara <i>cluster</i> dengan jenis	56
Gambar 4.14 Hubungan antara <i>cluster</i> dengan bulan.....	57
Gambar 4.15 Hubungan antara <i>cluster</i> dengan tahun.....	59

DAFTAR TABEL

Tabel 2.1 Daftar penelitian dengan metode dan topik terkait.....	6
Tabel 2.2 Langkah - langkah <i>K-means Clustering</i> (Vathy-Fogarassy & Abonyi, 2013)....	13
Tabel 2.3 Pemilihan Hipotesis Korelatif (Dahlan, 2011)	14
Tabel 2.4 Interpretasi Terhadap Koefisien Korelasi (Wirawan, 2016).....	15
Tabel 2.5 Tabel Kontingensi	17
Tabel 3.1 Atribut yang dihasilkan dari berkas spreadsheet	24
Tabel 3.2 Contoh nilai kosong pada data	27
Tabel 3.3 Contoh perubahan format penulisan angka desimal.....	28
Tabel 3.4 Daftar atribut dataset setelah <i>data transformation</i>	31
Tabel 3.5 Konstruksi atribut <i>hperkg</i> dari nilai atribut penjualan dan produksi.....	32
Tabel 3.6 Contoh hasil normalisasi atribut produksi dan atribut <i>hperkg</i>	33
Tabel 3.7 Contoh data yang digunakan untuk <i>clustering</i>	38
Tabel 3.8 Jarak data terhadap tiap <i>centroid</i>	39
Tabel 4.1 Nilai SSE terkecil untuk setiap <i>k</i>	46
Tabel 4.2 Informasi <i>cluster</i> yang terbentuk.....	48

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kabupaten Cilacap sebagai Kabupaten yang memiliki wilayah pantai merupakan salah satu daerah penghasil produk laut yang ada di Indonesia. Koperasi Unit Desa (KUD) Mino Saroyo sebagai koperasi nelayan melayani berbagai kebutuhan nelayan, salah satunya ialah pemasaran. KUD Mino Saroyo mengelola Tempat Pelelangan Ikan (TPI) yang digunakan nelayan dalam memasarkan hasil tangkapannya. Terdapat 9 unit TPI yang berada di masing-masing kelompok nelayan. TPI yang dikelola KUD Mino Saroyo mencatat setiap hasil lelang nelayan.

Data hasil tangkapan nelayan Kabupaten Cilacap disimpan dalam bentuk tabel. Data yang berupa tabel berisi data seperti jenis tangkapan, harga, serta berat tangkapan nelayan. Data yang dikumpulkan kemudian dilaporkan dalam bentuk jumlah setiap bulannya. Sampai saat ini data yang disimpan tersebut hanya digunakan sebagai laporan yang digunakan untuk menghitung detail pemasukan KUD Mino Saroyo. Data hasil tangkapan nelayan Kabupaten Cilacap yang telah terkumpul dapat dimanfaatkan untuk mencari informasi yang hanya bisa didapatkan dengan mengolah data tersebut. Salah satu cara untuk mencari informasi yang terdapat pada data adalah *data mining*.

Data mining merupakan proses yang melibatkan berbagai metode dengan tujuan menambang informasi yang bermanfaat dalam kumpulan data (Turban, et al., 2005). Dengan mengimplementasikan *data mining* terhadap data hasil tangkapan nelayan Kabupaten Cilacap maka selain menambah fungsi dari data, kegiatan tersebut juga bermanfaat untuk menggali informasi yang sebelumnya sulit diketahui hanya dengan menganalisa data mentah. Beberapa metode *data mining* yang dapat digunakan diantaranya adalah *K-d tree*, *Linear Regression*, *Association rules* dan lain lain (Witten, et al., 2017). Dalam tugas akhir ini, Data hasil tangkapan nelayan Kabupaten Cilacap dikelompokkan agar dapat diketahui makna dari hasil pengelompokan tersebut. Metode *data mining* yang digunakan adalah *Clustering*. *Clustering* digunakan dalam pengelompokan data hasil tangkapan nelayan Kabupaten Cilacap.

Clustering merupakan metode yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripannya. Terdapat dua pendekatan dalam melaksanakan *clustering*, *hierarchical* dan *partitonal*. Pada pendekatan *hierarchical* pengelompokan data dilaksanakan bertingkat, sehingga didapat berbagai hasil pembagian berdasarkan tingkatannya. Pada pendekatan *partitonal* pengelompokan data lebih berfokus terhadap hasil pembagian data. Pendekatan *partitonal* digunakan dalam tugas akhir ini dikarenakan jumlah data yang digunakan banyak, dimana pendekatan secara *hierarchical* terlalu memberi beban komputasi (Jain, et al., 1999). *Partitonal clustering* yang digunakan adalah *K-means Clustering*.

K-means Clustering merupakan salah satu metode yang dapat digunakan untuk mengelompokkan data berdasarkan kemiripan. Metode *K-means Clustering* juga salah satu metode *clustering* tertua yang masih sering digunakan. *K-means Clustering* termasuk dalam kategori *partitonal clustering*, dimana setiap *cluster* tidak memiliki anggota yang tumpang tindih keanggotaanya. Selain itu pada metode *K-means Clustering*, semua data yang ada memiliki *clusternya* masing masing (Wu, 2012). Dibandingkan dengan *partitonal clustering* lainnya seperti *K-medoid Clustering*, *K-means Clustering* memiliki kompleksitas yang lebih kecil (Velmurugan & Santhanam, 2010). Sebagai salah satu dari *unsupervised learning*, tujuan dari *clustering* salah satunya adalah memahami makna dari *cluster* yang terbentuk. *K-means Clustering* dipilih sebagai metode *clustering* dalam tugas akhir ini dikarenakan sifatnya yang memastikan setiap data memiliki *cluster* dan tidak tumpang tindih keanggotaanya sehingga pengamatan dapat terfokus pada *cluster* yang terbentuk.

Dalam analisis data, merepresentasikan data sebagai objek visual dapat membantu pengamat dalam mengetahui sifat dan kecenderungan dari data. Sifat manusia untuk membandingkan dan mencari pola pada objek menjadikan visualisasi data sebagai kegiatan yang dapat mempermudah proses analisis data (Fayyad, et al., 2002). Secara visual, data dalam bentuk *cluster* lebih mudah untuk dianalisis dibandingkan berupa list data.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah ditulis maka dapat disusun rumusan masalah yaitu :

1. Bagaimana mendapatkan hasil *clustering* data tangkapan nelayan Kabupaten Cilacap yang baik.
2. Bagaimana teknik visualisasi data yang digunakan untuk memvisualisasikan hasil *clustering* data tangkapan nelayan Kabupaten Cilacap.
3. Bagaimana menginterpretasikan hasil *clustering* data tangkapan nelayan yang ditampilkan secara visual.

1.3. Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah dibuat maka, tujuan dari tugas akhir ini adalah :

1. Membuat *cluster* dari data tangkapan nelayan Kabupaten Cilacap menggunakan *K-means Clustering* serta mengevaluasi hasil *cluster* yang dihasilkan.
2. Membuat visualisasi dari data *cluster* yang dihasilkan agar dapat memudahkan proses analisis.
3. Mengevaluasi pola yang dihasilkan dengan mengamati hasil visualisasi dan hubungan antar atribut data tangkapan nelayan Kabupaten Cilacap yang dikelompokkan menggunakan *K-means Clustering*. Dari hasil analisis tersebut diharapkan terdapat informasi yang sebelumnya belum diketahui dari data yang digunakan dalam tugas akhir ini.

Manfaat yang diharapkan dari tugas akhir ini adalah, hasil analisis pola tangkapan nelayan Kabupaten Cilacap dapat digunakan sebagai salah satu rujukan bagi KUD Mino Saroyo sebagai pengelola TPI untuk mengembangkan usahanya.

1.4. Ruang Lingkup

Ruang lingkup Pengelompokan Hasil Tangkapan Nelayan Kabupaten Cilacap adalah :

1. Data yang digunakan dalam tugas akhir ini dibatasi dengan data yang diperoleh dari hasil pelelangan dari TPI yang dikelola KUD Mino Saroyo Kabupaten Cilacap.
2. Data yang digunakan adalah data hasil tangkapan nelayan Kabupaten cilacap dari tahun 2015 sampai dengan 2017.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan dalam pembuatan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan tinjauan pustaka berupa dasar teori dan penelitian yang terkait dengan judul tugas akhir. Dasar teori dan penelitian terkait seperti teori tentang data mining dan penelitian yang membahas tentang data hasil laut digunakan sebagai landasan bagaimana tugas akhir ini dikerjakan.

BAB III METODOLOGI PENELITIAN

Bab ini membahas metode yang digunakan dalam melaksanakan tugas akhir. Dalam tahapan penyelesaian masalah, dijelaskan setiap tahap yang dilakukan seperti tahapan *data preprocessing*, *data mining*, *patern evaluation* dan *knowledge presentation*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari tugas akhir yang dilaksanakan serta pembahasannya. Hasil dari tugas akhir ini berupa pengujian *clustering*, pengujian hubungan antar data, dan hasil analisis yang didapat dari visualisasi data.

BAB V PENUTUP

Bab ini berisi kesimpulan apa saja yang dapat diambil dari hasil pengelompokan data nelayan Kabupaten Cilacap menggunakan *K-means Clustering* beserta saran untuk pengembangan penelitian lebih lanjut.